Package 'ENmix'

October 9, 2015

Version 1.0.0

Date 2015-04-15

Title Data preprocessing and quality control for Illumina HumanMethylation450 BeadChip

Type Package

Description Illumina HumanMethylation450 BeadChip has a complex array design, and the measurement is subject to experimental variations. The ENmix R package provides tools for low level data preprocessing to improve data quality. It incorporates a model based background correction method ENmix, and provides functions for inter-array quantile normalization, data quality checking, exploration of multimodally distributed CpGs and source of data variation. To support large scale data analysis, the package also provides multi-processor parallel computing wrappers for some commonly used data preprocessing methods, such as BMIQ probe design type bias correction and ComBat batch effect correction.

Depends minfi,parallel,doParallel,Biobase (>= 2.17.8),foreach

Imports MASS,preprocessCore,wateRmelon,sva,geneplotter

Suggests minfiData (>= 0.4.1), RPMM, RUnit, BiocGenerics

biocViews DNAMethylation, Preprocessing, QualityControl, TwoChannel, Microarray, OneChannel, MethylationArray, BatchEffect, Normalization, DataImport

License Artistic-2.0

NeedsCompilation no

Author Zongli Xu [cre, aut],
Liang Niu [aut],
Leping Li [ctb],
Jack Taylor [ctb]

Maintainer Zongli Xu <xuz@niehs.nih.gov>

R topics documented:

`bmiq.mc` *A multi-processor wrapper of BMIQ method*

Description

A multi-processor wrapper of BMIQ method. BMIQ is an intra-sample normalization procedure to correct the bias of Infinium 2 probe methylation beta values.

Usage

```
bmiq.mc(mdat, nCores = 1,...)
```

Arguments

`mdat`	An object of class `MethylSet`.
`nCores`	Number of cores used for computation.
`...`	See BMIQ in R package wateRmelon for more options.

Value

A data matrix of Methylation beta value.

Author(s)

Zongli Xu

References

Teschendorff AE et. al. *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data*. Bioinformatics. 2013

See Also

See `BMIQ` in R package `watermelon` for model details

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=normalize.quantile.450k(mdat,method="quantile1")
beta=bmiq.mc(mdatq1,nCores=10)
}}
```

`ComBat.mc` *A multi-processor wrapper for ComBat method.*

Description

A multi-processor wrapper for ComBat method. ComBat is a method to adjust batch effect where the batch covariate is known.

Usage

```
ComBat.mc(dat,batch,nCores = 1,...)
```

Arguments

`dat`	A data matrix with column for samples and row for probe.
`batch`	Batch covariate (multiple batches allowed)
`nCores`	Number of cores will be used for computation
`...`	See `ComBat` in `sva` package for extra options

Value

A data matrix with the same dimension as input data, adjusted for batch effects.

Author(s)

Zongli Xu

References

Johnson, WE, Rabinovic, A, and Li, C (2007). *Adjusting batch effects in microarray expression data using Empirical Bayes methods. Biostatistics 8(1):118-127.*

See Also

See `ComBat` in `sva` package for details.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdat=normalize.quantile.450k(mdat,method="quantile1")
beta=bmiq.mc(mdat,nCores=10)
batch=factor(pData(mdat)$Slide)
betaC=ComBat.mc(beta,batch,nCores=6,mod=NULL)
}}
```

---

nmode.mc *Estimating number of mode in methylaion data for each probe.*

Description

Due to SNPs in CpG probe region or other unknow factors, methylation beta values for some CpGs have multimodal distribution. This function is to identify this type of probes with obovious multimoal distribution.

Usage

```
nmode.mc(x, minN = 3, modedist=0.2, nCores = 1)
```

Arguments

x	A methylation beta value matrix with row for probes and column for samples.
minN	Minimum number of data points at each cluster
modedist	Minimum mode distance
nCores	Number of cores used for computation

Details

This function used an empirical approach to estimate number of mode in methylation beta value for each CpG probe. By default, the function requires the distance between modes have to be greater than 0.2 in methylation beta value, and each mode clusters should has at least 3 data points or 5% of data points whichever is greater.

Value

A vector of integers

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Under review

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
nmode=nmode.mc(beta, minN = 3,modedist=0.2, nCores = 5)
}}
```

---

normalize.quantile.450k

Quantile normalization.

Description

Quantile normalization of methylation intensity data across samples for Illumina Infinium HumanMethylation450 BeadChip.

Usage

```
normalize.quantile.450k(mdat, method = "quantile1")
```

Arguments

mdat	An object of class MethylSet.
method	Quantile normalization method. This should be one of the following strings: "quantile1", "quantile2", or "quantile3".

Details

By default, method = "quantile1" will separately quantile normalize Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile2" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile3" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I and II probes together.

Value

An object of class MethylSet.

Author(s)

Zongli Xu

References

Pidsley, R., CC, Y.W., Volta, M., Lunnon, K., Mill, J. and Schalkwyk, L.C. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics, 14, 293.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=normalize.quantile.450k(mdat,method="quantile1")
}}
```

---

pcrplot *Principal component regression plot*

Description

Principal component regression to explore methylation data variance structure and identifying possible confounding covariates. Principal components are derived using singular value decomposition method in beta value matrix.

Usage

```
pcrplot(beta, cov,npc=50)
```

Arguments

beta	A methylation beta value matrix with row for probes and column for samples.
cov	A data frame of covariates. Categorical variables should be converted to factors.
npc	The number of top principal components to plot

Value

A jpeg figure "svdscreeplot.jpg" to show the variations explained by each principal component.

A jpeg figure "pcr_diag.jpg" to show association strength between principal components and covariates with cell colors indicating different levels of association P values.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Under review

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
group=pData(mdat)$Sample_Group
slide=factor(pData(mdat)$Slide)
cov=data.frame(group,slide)
pcrplot(beta,cov,npc=6)
}}
```

`plotCtrl` *Plot internal controls of 450K BeadChip.*

Description

Intensity data are ploted for all internal control probe types on the Illumina Infinium HumanMethylation450 BeadChip. These figures can be used to check data quality and experimental procedures.

Usage

```
plotCtrl(rgSet)
```

Arguments

`rgSet`	An object of class `RGChannelSet`.

Value

A set of jpeg figures.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Under review.

Examples

```
if(FALSE){
if (require(minfiData)) {
plotCtrl(RGsetEx)
}}
```

`preprocessENmix` *The ENmix background correction for HumanMethylation450k BeadChip*

Description

ENmix models methylation signal intensities with a flexible exponential-normal mixture distribution, and models background noise with a truncated normal distribution. ENmix will split 450k BeadChip intensity data into 6 parts and separately model methylated and unmethylated intensities, 2 different color channels and 2 different probe designs.

Usage

```
preprocessENmix(rgSet, bgParaEst = "oob", dyeCorr=FALSE, nCores = 2)
```

Arguments

`rgSet`	An object of class `RGChannelSetExtended`, `RGChannelSet` or `MethylSet`.
`bgParaEst`	Optional method to estimate background normal distribution parameters. This must be one of the strings: "oob","est", or "neg".
`dyeCorr`	Dye bias correction: "TRUE" or "FALSE"
`nCores`	Number of cores will be used for computation

Details

By default, ENmix will use out-of-band Infinium I intensities ("oob") to estimate normal distribution parameters to model background noise. Option "est" will use combined methylated and unmethylated intensities to estimate background distribution parameters separately for each color channel and each probe type. Option "neg" will use 600 chip internal controls probes to estimate background distribution parameters. If rgSet if a `MethylSet`, then only option "est" can be selected.

Value

An object of class `MethylSet`

Author(s)

Zongli Xu and Liang Niu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. In review.

See Also

Package `minfi` for classes `RGChannelSet` and `MethylSet`

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
}}
```

`QCfilter` *Filter samples and probes with low data quality.*

Description

Filter samples and probes with low data quality measured based on detection P values and number of beads.

Usage

```
QCfilter(mdat, qcinfo, detPthre = 0.05, nbthre = 3,
samplethre = 0.01, CpGthre = 0.05,bisulthre=1,plot=FALSE,
        outid=NULL, outCpG=NULL)
```

Arguments

`mdat`	An object of class `MethylSet`
`qcinfo`	An object outputted from `QCinfo`
`detPthre`	Detection P value threshold
`nbthre`	Number of bead threshold
`samplethre`	Threshold for filtering samples, the percentage of low quality methylation data points across probes for each sample
`CpGthre`	Threshold for filtering probes, for percentage of low quality methylation data points across samples for each probe
`bisulthre`	Threshold for bisulfite to exclude samples.
`plot`	Whether to plot total intensity and beta value density plot before and after QC.
`outid`	User specified sample list to be filtered out. Valid ids are selected from colnames(mdat).
`outCpG`	User specified CpG list to be filtered out. Valid ids are selected from rownames(mdat).

Value

An object of class `MethylSet` with low data quality samples and probes filtered.

If option plot=TRUE, four figures will be produced: density_total_intensity_beforeQC.jpg, density_total_beta_beforeQC.jpg, density_total_intensity_afterQC.jpg, density_total_beta_afterQC.jpg.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Under review.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.450k.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.450k.exp(targets = sheet,extended = TRUE)
qcscore<-QCinfo(rgSet)
mdat=preprocessENmix(rgSet,bgParaEst="oob",nCores=6)
#filter low quality samples and probes
mdat1=QCfilter(mdat,qcinfo=qcscore)
}}
```

QCinfo *QC information.*

Description

Extract informations for data quanlity controls: detection P values and number of beads for each call of methylation beta value.

Usage

```
QCinfo(rgSet)
```

Arguments

rgSet	An object of class RGChannelSetExtended.

Value

A list with 2 data matrices (for detection P value and number of beads) and a numeric array (average intensities for bisulfite conversion controls probes).

Figure "qc_sample_1.jpg": scatter plot for Percent of low quality data per sample and Average bisulfite conversion intensity

Figure "qc_sample_2.jpg": histogram for Percent of low quality data per sample and Average bisulfite conversion intensity

Figure "qc_CpG.jpg": histogram for Percent of low quality data per CpG.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Under review.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.450k.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.450k.exp(targets = sheet,extended = TRUE)
qcscore<-QCinfo(rgSet)
}}
```

Index